Available online at www.sciencedirect.com
ScienceDirect
J. Genet. Genomics 36 (2009) 267–276
JOURNAL OF GENETICS AND GENOMICS
www.jgenetgenomics.org

Molecular analysis of rice plants harboring a multi-functional T-DNA tagging system

Yimian Ma [a, b], Luo Liu [a], Chengguang Zhu [a], Changhui Sun [a], Bo Xu [a, b], Jun Fang [a], Jiuyou Tang [a, b], Anding Luo [a], Shouyun Cao [a], Gupo Li [a], Qian Qian [c], Yongbiao Xue [d], Chengcai Chu [a, *]

[a] *State Key Laboratory of Plant Genomics and National Plant Gene Research Center (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*
[b] *Graduate University of the Chinese Academy of Sciences, Beijing 100039, China*
[c] *State Key Laboratory of Rice Biology, China Institute of Rice Research, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China*
[d] *Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*

Received for publication 10 December 2008; revised 23 February 2009; accepted 26 February 2009

Abstract

About 25,000 rice T-DNA insertional mutant lines were generated using the vector pCAS04 which has both promoter-trapping and activation-tagging function. Southern blot analysis revealed that about 40% of these mutants were single copy integration and the average T-DNA insertion number was 2.28. By extensive phenotyping in the field, quite a number of agronomically important mutants were obtained. Histochemical GUS assay with 4,310 primary mutants revealed that the GUS-staining frequency was higher than that of the previous reports in various tissues and especially high in flowers. The T-DNA flanking sequences of some mutants were isolated and the T-DNA insertion sites were mapped to the rice genome. The flanking sequence analysis demonstrated the different integration pattern of the right border and left border into rice genome. Compared with *Arabidopsis* and poplar, it is much varied in the T-DNA border junctions in rice.

Keywords: rice; mutant population; promoter trap; activation tag; T-DNA integration

Introduction

Rice (*Oryza sativa* L.) is one of the most important cereal species of the world. It has the smallest genome in monocot plants and supplies more than 50% calories for the world population. Since the rice genome sequencing project was first initiated by the publicly funded International Rice Genome Sequencing Project (IRGSP) in 1997, two draft maps of different subspecies of rice were published in 2002 by Syngenta (Torrey Mesa Research Institute, San Diego, USA) and BGI (Beijing Genomics Institute, Beijing, China) (Goff et al., 2002; Yu et al., 2002). In 2004, the IRGSP declared the completion of the japonica rice genome with high-accuracy, and the finished rice genome sequence covered ~95% of the 389 Mb of the rice genome with an error rate of less than one per 10 kb. At the same time, the big progress has also been made in rice functional genome research with large scale rice T-DNA insertional mutant population generated in different laboratories of many countries, such as Korea (Jeon et al., 2000; Jeong et al., 2002; Ryu et al., 2004; Jeong et al., 2006), England (Johnson et al. 2005), and France (Sallaud et al.

* Corresponding author. Tel & Fax: +86-10-6487 7570.
E-mail address: ccchu@genetics.ac.cn

DOI: 10.1016/S1673-8527(08)60114-9

2003, 2004). In China, several large rice mutant pools had been developed in Shanghai, Wuhan (Wu et al., 2003; Zhang et al., 2006), Beijing (Wan et al., 2009), and Taiwan (Chern et al., 2007; Hsing et al., 2007). It was known that the total number of the rice T-DNA insertional mutants reached more than 425,000 worldwide, with some mutant lines and flanking sequences tags (FSTs) accessible on website (http://ship.plantsignal.cn/home.do; http://rmd.ncpgr.cn/; http://trim.sinica.edu.tw/; http://www.postech.ac.kr/life/pfg/risd). Gene trap and activation tag are two methods frequently used in T-DNA insertional mutagenesis studies. Gene trap system provides a way of identifying novel genes based on their expression patterns. Since the first gene trap system was established (Casadaban and Cohen, 1979), it has been widely used in gene function analysis both in animals and in plants. In plants, many important genes were cloned by this way (Claes et al., 1991; Lindsey et al., 1993, 1998; Topping et al., 1994; Wei et al., 1997; Topping and Lindsey, 1997; He and Gan, 2001). Activation tagging is a method to generate dominant mutations in plants or plant cells by random insertion of a T-DNA carrying constitutive enhancer elements, which can cause transcriptional activation of flanking plant genes (Memelink, 2003). Since the first establishment of activation tag in plants (Odell et al., 1985), functional genes were frequently isolated (Hayashi et al., 1992; Kardailsky et al., 1999; Borevitz et al., 2000; van der Graaff et al., 2000; van der Graaff et al., 2002; Busov et al., 2003; Nakazawa et al., 2003; Mathews et al., 2003). Because gene trap and activation tag have the advantage of high efficiency, the two methods are frequently used in the construction of multiple-functional T-DNA insertional mutant population. T-DNA in binary vector could be modified with either reporter gene near the border for gene trapping, or enhancer element as activation tag, and the reporter gene and enhancer element could also be combined to construct a multiple-function vector, such as pGA2715 and pGA2772, which were constructed at Pohang University of Science and Technology in Korea and were used to generate more than 34,000 rice mutant lines (Jeong et al., 2002, 2006). In this study, a binary vector combined with promoter trap and activation tag functions were also constructed and a rice mutant population with about 25,000 transformants was generated. A large-scale phenotyping in the field has been carried out, 471 flanking sequencing were made and the T-DNA integration sites were mapped to the rice genome. Based on these data, the distribution of T-DNA insertion sites and T-DNA integration patterns were also investigated.

Materials and methods

Construction of binary vector pCAS04

The vector construction was performed following standard protocols as described by Sambrook et al. (1989). The multifunctional vector pCAS04, contains the promoterless reporter β-glucuronidase gene (*GUS*) near to the right border and the rice *Actin1* promoter near to the left border, is derived from the binary vector pEP200. At first, the portion containing promoterless *GUS* gene and CaMV terminator was assembled into a pBluescript SK (−) vector, then cut by *Hin*d III and *Bam*H I and inserted in pEP200, resulting in the plasmid pEP200-GUS. Afterwards, the complete cassette of maize ubiquitin promoter:*nptII*:CaMV terminator from plasmid pJFNPTII (a kind gift from J. Xu, Institute of Plant Genetics and Crop Plant Research, Germany) was cut by restriction enzymes *Bst*X V and *Kpn* I and then inserted into the pEP200-GUS resulting in a plasmid pCAS03. At last, the rice *Actin 1* promoter was amplified and inserted into the binary vector pCAS03. The final plasmid pCAS04 was made after examination of the insert orientation by PCR.

Agrobacterium-mediated transformation of rice

Rice (*Oryza sativa* L. cv. *Nipponbare*) was used for transformation and *Agrobacterium*-mediated transformation was used as described by Liu et al. (2007).

Genomic DNA extraction and Southern blot

Genomic DNA was prepared from mature leaves or seedlings using SDS method with some modification. For identifying positive lines, a pair of primers for amplifying gene *nptII* was designed with Forward primer: TCCGGTG CCCTGAATGAACT, and Reverse primer: GGCGATAC CGTAAAGCACGA. For Southern blot, about 10 μg of purified DNA were digested with *Bam*H I, size-fractionated on a 0.8% agarose gel, and transferred onto a nylon membrane. The blots were then hybridized to *GUS* specific probe in hybridization buffer containing 0.5 mol/L sodium phosphate (pH 7.2), 7% SDS, 1% BSA, and 20 mmol/L EDTA. Final washes of the filters were carried out with wash solution (0.1 × SSC and 0.1% SDS) for 15 min at 65°C.

Histochemical GUS assay

Histochemical GUS assay of plant tissues was carried out as described by Jefferson et al. (1987). Leaves, flowers and roots, which had been hand-cut with a razor, were incubated at 37°C in 50 mmol/L phosphate buffer (pH 7.0) containing 1 mmol/L 5-bromo-4-chloro-3-indolyl-β-glucuronide (x-gluc), 10mmol/L mercaptoethanol, 10 mmol/L EDTA, 0.1% sodium lauryl sarcosine, and 0.1% Triton X-100 overnight, then washed in 70% ethanol to get rid of the background color of tissues.

Thermal asymmetric interlaced-PCR (TAIL-PCR) and sequencing analysis

TAIL-PCR was performed according to the standard protocol of Liu and Whittier (1995). The following nested primers complementary to the sequences of T-DNA borders were designed: RB1, TGCATCGGCGAACTGATCG T; RB2, CGCTTTCCCACCAACGCTGA; RB3, ACGGG TTGGGGTTTCTACAGG A; LB1, ACAATGCTGAGGG ATTCAAATTCTACCAC; LB2, TACTGAATTAACGCC GAATTGAATTCGAG. Four arbitrary degenerate primers were used: AD1, NTCGA(G/C)T(A/T)T(G/C)G(A/T)GTT; AD2, NGACGA (G/C)(A/T)G ANA (A/T)GAA; AD3, NGACGA(G/C)(A/T)GANA(A/T)GTT; AD4, NGACGA (G/C)(A/T)GANA(A/T)GAC.

PCR products were sequenced directly using the ABI PRISM 3730 DNA Analyzer System (ABI, USA). The flanking sequences were then BLAST against rice genome in the Rice Annotation Project Database (RAP-DB) (http://rapdb.dna.affrc.go.jp). Results with high similarity of E value<0.01 were used.

Northern hybridization and single-strand hybridization

Total RNA isolation and Northern hybridization method was described by Wu et al. (2004). The single-stranded RNA probes were transcribed from PCR products that amplified from linearized pSK(−) DNA plasmid using T7 RNA polymerase (Promega, USA). The filters were washed in 0.5 × SSC with 0.1% SDS for 30 min, followed by washing in 0.1 × SSC with 0.1% SDS for 30 min at 68°C after hybridization.

Results

Generation of rice mutant population

In this study, we constructed a binary vector pCAS04, which the *Actin1* promoter located to the left border as activation tagging, the promoterless *GUS* gene located to the right border as promoter trapping, and the selective marker gene was *npt*II (Fig. 1). About 25,000 independent T-DNA lines were generated by *Agrobacterium*-mediated transformation. Rice genomic DNA was isolated from about 3,000 T_0 transgenic lines for PCR analysis. Southern blot was performed for estimating the average copy number of integrated T-DNA. Genomic DNA from 50 randomly-selected T_0 mutants were digested with *Bam*H I and hybridized with a *GUS*-specific probe, about 40% of the transformants contained single-copy T-DNA insertion and the average copy number of T-DNA was 2.28 (Fig. 2). A total of more than 16,000 mutants in the field were screened out and many mutants with agronomically important traits were obtained (Fig. 3). Furthermore, the segregations of quite a few T_1 mutants fit the 3:1 Mendelian ratio. The phenotype of some mutants has been linked with the T-DNA inserts through PCR detection, but the co-segregation of T-DNA inserts and phenotypes was not such frequently detected as expectation.

Mapping T-DNA insertion sites to rice genome

We isolated the T-DNA flanking sequences tags (FSTs)

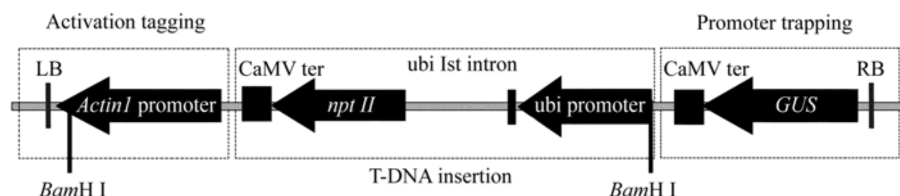

Fig. 1. A schematic diagram of the T-DNA region from the binary vector pCAS04. RB: the right border of T-DNA; LB: the left border of T-DNA; ubi: maize Ubiquitin promoter; *nptII*: neomycin phosphotransferase gene; CaMVter: CaMV terminator; *GUS*: β-glucuronidase gene; *Actin1*: rice *Actin1* promoter; ubi Ist intron: the first intron of ubiquitin promoter.

of some T-DNA inserts *via* TAIL-PCR method. Flanking sequences of about 1,000 T-DNA inserts were obtained from the 3,000 T_0 generation T-DNA transgenic lines. After directly sequencing of PCR products and BLAST against the RAP-DB database (http://rapdb.dna.affrc.go.jp) with the new released rice genome (IRGSP), 471 sequences were matched to rice genome (E<0.01) and the corresponding T-DNA insertion sites were mapped. The distribution feature of these insertion sites was classified as intergenic regions, 5′ UTR (Up-stream 1,000 bp), 3′ UTR (Down-stream 500 bp), exon and intron. Statistics of the number of various features was listed in Table 1. According to the *japonica* rice genome data released by IRGSP in 2005 (IRGSP, 2005), we supposed that the protein encoding regions with the upstream 1,000 bp and the downstream 500 bp are genic regions, and the ratio of the genic regions in the rice genome was 42.52%. According to the data in Table 1, the T-DNA insertions in rice intergenic region had a frequency of 55.63 % (262/471), so the correspondingly insertion ratio in genic area was 44.37%., which was more than the proportion of the genic regions in rice whole genome (42.52%). So it seems that there is a small shift to genic regions of these T-DNA inserts. We also estimated the T-DNA inserts preference in different part of a gene with the data acquired, it was indicated T-DNA inserts into rice genome with obvious bias towards 5′UTR and 3′UTR gene regulation regions so as to make gene knockout mutations. The genes T-DNA disrupted can be involved in various biological processes and molecular functions, most of them were annotated and some of them were hypothetical genes. Although the appearance of most of the 471 mutant lines had no obviously change, about 10% of them were dwarf or sterility in the maturation phase.

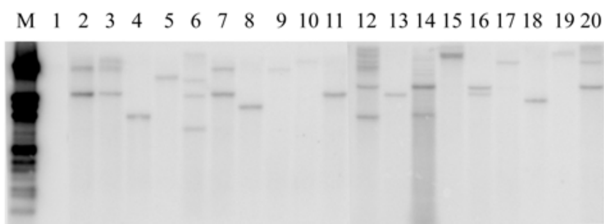

Fig. 2. Southern blot analysis of independent mutant lines. M: λ/*Eco*R I+*Hin*d III marker; 1: blank control; 2–20: represent different mutants.

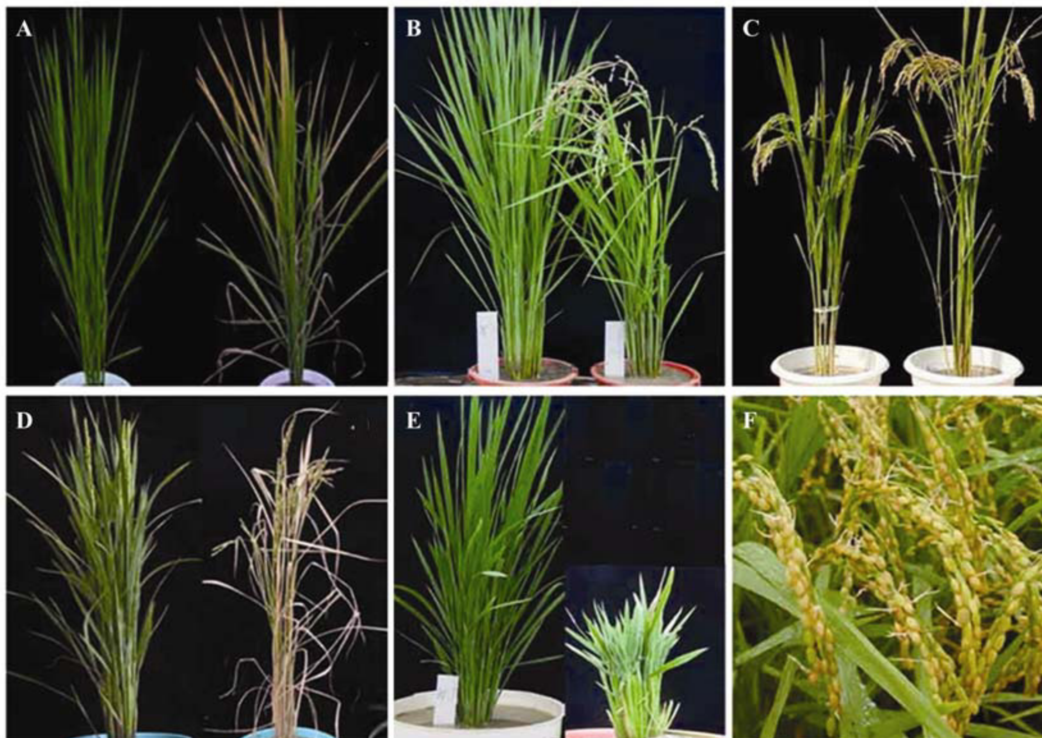

Fig. 3. T-DNA insertional mutants with altered agronomically-important phenotypes. **A:** high phosphate accumulating mutant; **B:** early flowering; **C:** elongated uppermost internode (*eui*); **D:** premature leaf senescence 2; **E:** dwarf; **F:** preharvest sprouting.

Table 1
Distribution of T-DNA insertion site in different chromosomes

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	Total
intergenic	37	40	32	27	14	18	23	14	12	19	18	8	262
Exon	11	2	11	4	4	2	4	1	1	2	2	3	47
Intron	9	6	8	4	4	4	3	4	4	3	2	2	53
5′UTR	16	11	15	6	2	5	1	6	1	2	3	4	72
3′UTR	9	5	7	1	1	2	3	2	0	3	2	2	37
Total	82	64	73	42	25	31	34	27	18	29	27	19	471

Analysis of the efficiency of promoter trapping

To test the efficiency of the promoter trapping system, GUS assay was carried out with leaves, roots, flowers, and immature seeds from about 4,310 primary transgenic rice plants. The histochemical data revealed that the efficiency of GUS staining was 4.4% (187/4,310) for leaves, 6.8% (255/3,729) for roots, 35.5% (134/391) for flowers and 8.08% (41/507) for immature seeds. In these GUS positive lines, 115 (29.4%) among 391 lines were flower-specific, 18 (3.5%) among 507 lines showed seed-specific expression pattern, 74 (1.7%) out of 4,310 leaf-specific, and 119 (3.2%) among 3,729 lines were root-specific (Table 2). Additionally, out of 391 flower-specific transgenic lines, 115 expression patterns were varied, such as pollen-specific expression, carpel-development-specific expression etc (Fig. 4).

Table 2
Preference of GUS expression in transgenic rice

Organ	Total	Stained lines	Ratio (%)	Preferential	Ratio (%)
Roots	3,729	255	6.8	119	3.2
Leaves	4,310	187	4.3	74	1.7
Flowers	391	134	35.5	115	29.4
Immature seeds	507	41	8.1	18	3.5

Cloning salt tolerant genes via activation tagging

We tested the activation tagging effect for 300 salt tolerant T_0 mutants. Twenty-nine T-DNA insertion sites were located on rice genome by TAIL-PCR and sequence BLAST. Downstream and upstream 10 kb of these T-DNA insertion sites were searched for the probably activation genes by *Actin1* promoter. The candidate genes encode sigma factor, putative trehalose-6-phosphate phosphatase, Myb-like protein, AP2 domain containing protein, Zf-MYND domain containing protein, and several hypothetical proteins, and these candidate genes were further tested *via* RT-PCR and Northern examination. A gene named *ST12* was obviously activated, and the full-length cDNA was cloned *via* RACE. Retransformation of this gene to rice could lead to salt tolerance. Another candidate gene encoded a putative trehalose-6-phosphate phosphatase, which showed low expression level in normal conditions, but induced by salt stress. The expression level of this gene in mutant was higher than WT seedling after salt stress for 6 hours, which suggest the gene enhanced expression under salt treated condition without alteration of its expression pattern. These results indicate our system can be used for gain-of-function analysis.

Analysis of T-DNA integration junctions in rice

To analyze the integration patterns of T-DNA in rice genome we have aligned the FSTs with sequence near T-DNA border. A total of 60 fragments flanking T-DNA right border (RB), and 60 fragments flanking left border (LB) with a good T-DNA border sequence definition were selected from 471 FSTs. The aligned sequence results revealed that T-DNA border had various ranges of deletions. In quite a few cases some fragments had no sequence homology with either the vector sequences nearby the T-DNA border or the rice genome, which were named as filler DNAs. In contrast, some nucleotides in the junction sites had sequence homology with nearby sequences at both the left and right sides, which were named as microhomologies (micro homology nucleotides). The sum of linkage with border deletion, filler DNAs, and microhomologies, together with the border deletion range, lengths of filler DNAs and microhomologies were listed in Table 3. From statistical data, it was shown that the T-DNA border deletions were more serious in the left border than right border. Furthermore, microhomologies of 1−11 base pairs range existed in more than 50% LB junctions and filler DNAs existed in about 80% RB junctions. The origin of

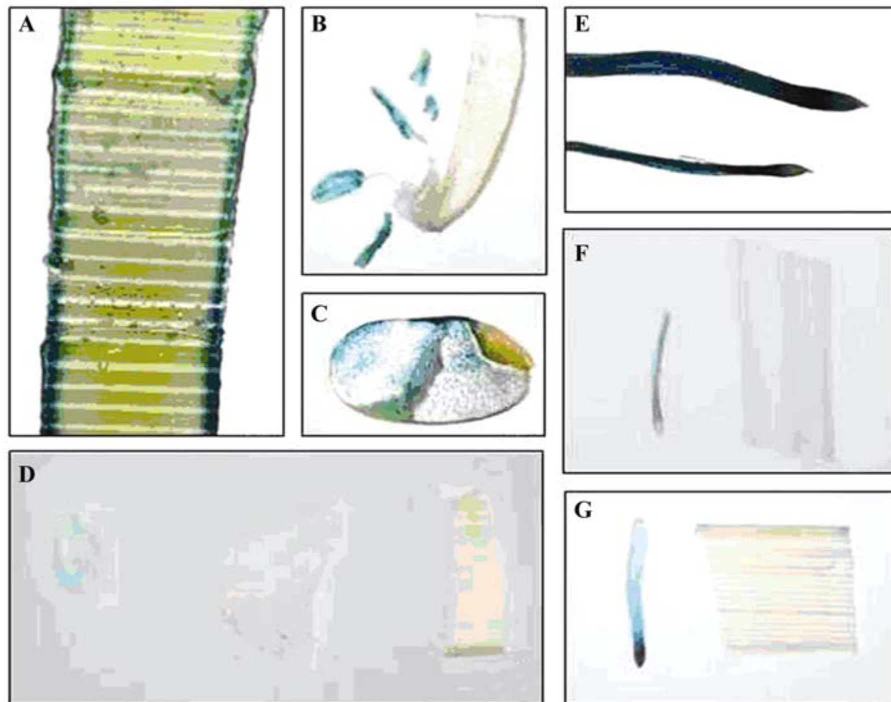

Fig. 4. GUS expression patterns in rice promoter trapping lines. **A:** specific expression in the leaves. **B:** specific expression in anther. **C:** specific expression in endosperm. **D:** specific expression in embryo. **E, F, G:** specific expression in different parts of roots.

Table 3
Different integration patterns of T-DNA border

Border	Left border	Right border
Deletions (range)	59 (3–147 bp)	56 (1–23 bp)
filler DNAs (length)	14 (4–55 bp)	50 (2–695 bp)
Microhomologies (length)	32 (1–11 bp)	5 (1–5 bp)

the large amount of filler DNA was also surveyed, and it was proposed to be derived from non-T-DNA vector sequences in the plasmid. To identify the T-DNA recombination hot sites at sequence near borders, we also made a detailed table of various linkage types based on the different T-DNA border deletion ranges (Supplemental Tables 1 and 2). For the right border, only 8 linkage types were identified, the T-DNA integrated hot sites were 12 bp away from the VirD protein cut site. For the left border, 38 different linkage types were found, and the T-DNA integration site with the most high frequency was 8 bp away from the VirD protein cut site. The analysis of the sequences character near the recombination hot sites were underway.

Discussion

We have generated a multiple-function rice mutant population with about 25,000 individuals, which is a valuable resource for rice functional genomic research. The multiple-function binary vector carries both promoter trapping and activation tagging, which is very similar to the vector pGA2715 and pGA2772 developed at POSTECH (Jeong et al., 2002, 2006). The main difference from the two vectors is that we used rice *Actin1* promoter as activation tagging in pCAS04 but not multimerized transcriptional enhancers from the CaMV 35S promoter. The experiment results showed the *Actin1* enhancer would have the same function in gene activation. In the mutant population, about 40% of the transformants contained single-copy T-DNA insertion and the average copy number of T-DNA was 2.28. Based on that, about 57,000 T-DNA inserts and about 10,000 single copy T-DNA insertion lines have been acquired in our mutant population. By extensive phenotyping in the field, quite a number of agronomically important mutants have been isolated. Despite the usefulness of a phenotype-driven genetic approach, it is somewhat inconvenient for high-throughput screening of rice mutants. Firstly, screening rice mutants requires more effort because of the larger size and longer life cycles. Secondly, the phenotypic alternations observed may not necessary due to insertional mutation, but to the transposition of endogenous mobile elements such as Tos17 or muta-

tions caused by tissue culture (Jeong et al., 2006). In addition, many mutants with serious change of appearance would have fewer seeds. This indicates many of the alterations of rice appearance may be a consequence of serious disruption of genome, such as genome deletion, chromosome reduplication or recombination, which results in a disruption of plant propagation. All of the reasons give rise to the difficulties of the phenotype-driven genetic approach, so reverse genetics approaches such as TAIL-PCR is applied in assistant of phenotype screening for functional genes isolation.

So far we have isolated 471 FSTs from these T-DNA mutants, and the corresponding 471 T-DNA insertion sites in rice genome were located. Although the shift to genic regions was not significant over these samples, the shift to the gene regulation regions was obvious. These results were consistent with previous analysis (Chen et al., 2003; Zhang et al., 2007).

Histochemical assay with our mutant population revealed that a very high GUS expression efficiency was found in various tissues with the highest expression in flowers (35.5%). In rice, Chin et al. (1999) used DS carrying a gene trap system, and found that GUS expression was detected in 8% of the lines in various parts of rice panicles. Then Ryu et al. (2004) discovered a 2.5% promoter trapping efficiency with GFP-positive. Compared with their studies, our promoter trapping lines showed a much higher expression level. The reason may be related to the *Actin1* enhancer element in the binary vector. In the research of Jeon et al. (2000), only 1.6%–2.1% of plants showed GUS activity in various tissues, but the GUS-staining frequency was increased about twice when an enhancer element was added. So it was suggested that the enhancer sequence present in the T-DNA improves the trapping efficiency, and the enhancers simply elevated the expression level of the nearby genes without altering the original expression pattern (Jeong et al., 2002). So the *Actin1* enhancer element may elevate the expression of genes nearby especially at flowers in our experiments. However, multiple copies of T-DNA may be inserted into one genome locus and can be result in tandem LB: RB. The tandem LB: RB may favor read through the RB, driven by the *Actin1* promoter which is very active in flowers, so the actual trapping efficiency may be lower than detected.

Previously, Ichikawa et al. (2003) have used ten activation tagging lines that showed phenotypes to monitor the expression levels of genes adjacent to the T-DNA integration sites in *Arabidopsis* activation tagging pool, activation was observed in 7 out of 17 of the adjacent genes detected (Ichikawa et al., 2003). In rice, among 15 selective lines for activation analysis in an activation tagging mutant pool, four showed the increasing expression level of the genes near the T-DNA insertion sites. In our analysis, we detected two activation tagging lines out of 18 candidate genes. The first gene *ST12*, showed a constitutive expression in mutant leaves, whereas a very low expression level detected in the leaves of wild type. The second gene, which encodes a putative trehalose-6-phosphate phosphatase, was induced by salt stress after 6 hours, the expression level increased in mutant under induced condition with the same expression pattern. Therefore, the *Actin1* promoter did elevate the gene expression level without change its expression pattern, which is similar to the CaMV35S enhancer. In our research, some genes were not activated even just at the downstream of *Actin1* promoter. This may result from the presence of insulator elements that prevent the misregulation of genes by restricting the effects of the regulatory elements to specific domains (Oki and Kamakaka, 2002).

In our studies, many microhomologies were discovered in LB junctions, and large amount of filler DNAs were discovered in RB junctions. This provided evidence for the different integration patterns between LB and RB in T-DNA mediated transformation in rice. With a survey of filler DNAs of RB, we propose the filler DNA should be originated from either the sequences nearby T-DNA border or non-T-DNA vector sequences. This is supported by lots of reports that non-T-DNA vector sequences remained in the T-DNA transformation process in dicot plants (Cluster et al., 1996; Kononov et al., 1997; Wenck et al., 1997; de Buck et al., 2000; bdal-Aziz et al., 2006) and also monocot plant rice (Kim et al., 2003). The most common integration mode of right border was12 bp away from the VirD cut site, and was 8 bp away from the VirD cut site for the left border. We further compared the T-DNA border deletion range with *Arabidopsis*, poplar (*Populus tremula*), and tomato (*Salanum lycopersicum* var. moneymaker), it was found that in *Arabidopsis* most integration sites retained a portion of the LB and the truncations ranged from 27 to 109 bp (Mayerhofer et al., 1991; Gheysen et al., 1991), in poplar, 1/5 of T-DNA integrations contained an intact LB and truncations varied between 2 and 24 bp (Kumar and Fladung, 2002), in tomato, 1 out of 42 contains an intact LB, all others did not contain a portion of LB, and deletions from the cleavage site can reach 400 bp (Thomas and Jones, 2007). In our studies in rice, only 1/60 contains an intact LB, the deletions from the cleavage site can reach 147 bp. In the case of RB, it is generally believed RB

transfer is more precise than LB transfer because of the protection of VirD2 protein. In *Arabidopsis* most integration contained a nearly complete RB repeat, but truncations can reach 34 bp (Mayerhofer et al., 1991; Gheysen et al., 1991). In poplar 19/27 T-DNA integrations contained a complete RB and truncations in the remainder varied between 2 and 23 bp (Kumar and Fladung, 2002). In tomato, 26/55 was located in 10 bp of the RB terminus, and 5/28 contained an intact RB repeat. However, the integration with complete RB repeat was only 1/15 in our investigation in rice and deletions can be 1–23 bp. So the integration in both LB and RB in rice was varied more than *Arabidopsis*, poplar, and also different from tomato. In fact, Kim et al. (2003) also indicated the border variation of rice is different from other species. This indicated T-DNA mediated recombination in difference species may be differed which probably caused by particular proteins of plant host participated in this process.

Acknowledgements

This work was kindly supported by the National High Technology Research and Development Program of China (No. 2002AAZ2001) and the National Natural Sciences Foundation of China (No. 30270758 and 30621001).

Supplemental data

Supplemental Tables 1 and 2 associated with the article can be found in the online version at www.jgenetgenomics.org.

References

bdal-Aziz, S.A., Pliego-Alfaro, F., Quesada, M.A., and Mercado, J.A. (2006). Evidence of frequent integration of non-T-DNA vector backbone sequences in transgenic strawberry plant. J. Biosci. Bioeng. **101:** 508–510.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12:** 2383–2394.

Busov, V.B., Meilan, R., Pearce, D.W., Ma, C., Rood, S.B., and Strauss, S.H. (2003). Activation tagging of a dominant gibberellin catabolism gene *(GA 2-oxidase*) from *poplar* that regulates tree stature. Plant Physiol. **132:** 1283–1291.

Casadaban, M.J. and Cohen, S.N. (1979). Lactose genes fused to exogenous promoters in one step using a Mu-*lac* bacteriophage: *In vivo* probe for transcriptional control sequences. Proc. Natl. Acad. Sci. USA **76:** 4530–4533.

Chen, S., Jin, W., Wang, M., Zhang, F., Zhou, J., Jia, Q., Wu, Y., Liu, F., and Wu, P. (2003). Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J. **36:** 105–113.

Chern, C.G., Fan, M.J., Yu, S.M., Hour, A.L., Lu, P.C., Lin, Y.C., Wei, F.J., Huang, S.C., Chen, S., Lai, M.H., Tseng, C.S., Yen, H.M., Jwo, W.S., Wu, C.C., Yang, T.L., Li, L.S., Kuo, Y.C., Li, S.M., Li, C.P., Wey, C.K., Trisiriroj, A., Lee, H.F., and Hsing, Y.I. (2007). A rice phenomics study--phenotype scoring and seed propagation of a T-DNA insertion-induced rice mutant population. Plant Mol. Biol. **65:** 427–438.

Chin, H.G., Choe, M.S., Lee, S.H., Park, S.H., Koo, J.C., Kim, N.Y., Lee, J.J., Oh, B.G., Yi, G.H., Kim, S.C., Choi, H.C., Cho, M.J., and Han, C.D. (1999). Molecular analysis of rice plants harboring an Ac/Ds transposable element-mediated gene trapping system. Plant J. **19:** 615-623.

Claes, B., Smalle, J., Dekeyser, R., van Montagu, M., and Caplan, A. (1991). Organ-dependent regulation of a plant promoter isolated from rice by 'promoter-trapping' in tobacco. Plant J. **1:** 15–26.

Cluster, P.D., O'Dell, M., Metzlaff, M., and Flavell, R.B. (1996). Details of T-DNA structural organization from a transgenic Petunia population exhibiting co-suppression. Plant Mol. Biol. **32:** 1197–1203.

de Buck, S., de Wilde, C., van Montagu, M., and Depicker, A. (2000). Determination of the T-DNA transfer and the T-DNA integration frequencies upon cocultivation of *Arabidopsis thaliana* root explants. Mol. Plant Microbe Interact. **13:** 658–665.

Gheysen, G., Villarroel, R., and van Montagu, M. (1991). Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev. **5:** 287–297.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002). A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica)*. Science **296:** 92-100.

Hayashi, H., Czaja, I., Lubenow, H., Schell, J., and Walden, R. (1992). Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. Science **258:** 1350–1353.

He, Y., and Gan, S. (2001). Identical promoter elements are involved in regulation of the *OPR1* gene by senescence and jasmonic acid in *Arabidopsis*. Plant Mol. Biol. **47:** 595–605.

Hsing, Y.I., Chern, C.G., Fan, M.J., Lu, P.C., Chen, K.T., Lo, S.F., Sun, P.K., Ho, S.L., Lee, K.W., Wang, Y.C., Huang, W.L., Ko, S.S., Chen, S., Chen, J.L., Chung, C.I., Lin, Y.C., Hour, A.L., Wang, Y.W., Chang, Y.C., Tsai, M.W., Lin, Y.S., Chen, Y.C., Yen, H.M., Li, C.P., Wey, C.K., Tseng, C.S., Lai, M.H., Huang, S.C., Chen, L.J., and Yu, S.M. (2007). A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol. Biol. **63:** 351–364.

Ichikawa, T., Nakazawa, M., Kawashima, M., Muto, S., Gohda, K., Suzuki, K., Ishikawa, A., Kobayashi, H., Yoshizumi, T., Tsumoto, Y., Tsuhara, Y., Iizumi, H., Goto, Y., and Matsui, M. (2003). Sequence database of 1172 T-DNA insertion sites in *Arabidopsis* activation-tagging lines that showed phenotypes in T1 generation. Plant J. **36:** 421–429.

IRGSP. (2005). The map-based sequence of the rice genome. Nature **436:** 793–800.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fu-

sions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6:** 3901–3907.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., Kim, C., Jang, S., Yang, K., Nam, J., An, K., Han, M.J., Sung, R.J., Choi, H.S., Yu, J.H., Choi, J.H., Cho, S.Y., Cha, S.S., Kim, S.I., and An, G. (2000). T-DNA insertional mutagenesis for functional genomics in rice. Plant J. **22:** 561–570.

Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K., and An, G. (2002). T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. **130:** 1636–1644.

Jeong, D.H., An, S., Park, S., Kang, H.G., Park, G.G., Kim, S.R., Sim, J., Kim, Y.O., Kim, M.K., Kim, S.R., Kim, J., Shin, M., Jung, M., and An, G. (2006). Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. Plant J. **45:** 123–132.

Johnson, A.A., Hibberd, J.M., Gay, C., Essah, P.A., Haseloff, J., Tester, M., and Guiderdoni, E. (2005). Spatial control of transgene expression in rice (*Oryza sativa* L.) using the GAL4 enhancer trapping system. Plant J. **41:** 779–789.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation tagging of the floral inducer *FT*. Science **286:** 1962–1965.

Kim, S.R., Lee, J., Jun, S.H., Park, S., Kang, H.G., Kwon, S., and An, G. (2003). Transgene structures in T-DNA-inserted rice plants. Plant Mol. Biol. **52:** 761–773.

Kononov, M.E., Bassuner, B., and Gelvin, S.B. (1997). Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. Plant J. **11:** 945–957.

Kumar, S., and Fladung, M. (2002). Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration. Plant J. **31:** 543–551.

Lindsey, K., Topping, J.F., Muskett, P.R., Wei, W., and Horne, K.L. (1998). Dissecting embryonic and seedling morphogenesis in *Arabidopsis* by promoter trap insertional mutagenesis. Symp. Soc. Exp. Biol. **51:** 1–10.

Lindsey, K., Wei, W., Clarke, M.C., McArdle, H.F., Rooke, L.M., and Topping, J.F. (1993). Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. Transgenic Res. **2:** 33–47.

Liu, X.Q., Bai, X.Q., Wang, X.J., and Chu, C.C. (2007). OsWRKY71, a rice transcription factor, is involved in rice defense response. J. Plant Physiol. **164**: 969-979.

Liu, Y.G., and Whittier, R.F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25:** 674–681.

Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., Schuster, D.K., Menasco, D.J., Wagoner, W., Lightner, J., and Wagner, D.R. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell **15:** 1689–1703.

Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K., Redei, G.P., Schell, J., Hohn, B., and Koncz, C. (1991). T-DNA integration: a mode of illegitimate recombination in plants. EMBO J. **10:** 697–704.

Memelink, J. (2003). T-DNA activation tagging. Methods Mol. Biol. **236:** 345–362.

Nakazawa, M., Ichikawa, T., Ishikawa, A., Kobayashi, H., Tsuhara, Y., Kawashima, M., Suzuki, K., Muto, S., and Matsui, M. (2003). Activation tagging, a novel tool to dissect the functions of a gene family. Plant J. **34:** 741–750.

Odell, J.T., Nagy, F., and Chua, N.H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313:** 810–812.

Oki, M., and Kamakaka, R.T. (2002). Blockers and barriers to transcription: competing activities? Curr. Opin. Cell Biol. **14:** 299–304.

Ryu, C.H., You, J.H., Kang, H.G., Hur, J., Kim, Y.H., Han, M.J., An, K., Chung, B.C., Lee, C.H., and An, G. (2004). Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. Plant Mol. Biol. **54:** 489–502.

Sallaud, C., Gay, C., Larmande, P., Bes, M., Piffanelli, P., Piegu, B., Droc, G., Regad, F., Bourgeois, E., Meynard, D., Perin, C., Sabau, X., Ghesquiere, A., Glaszmann, J.C., Delseny, M., and Guiderdoni, E. (2004). High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. Plant J. **39:** 450–464.

Sallaud, C., Meynard, D., van Boxtel, J., Gay, C., Bes, M., Brizard, J.P., Larmande, P., Ortega, D., Raynal, M., Portefaix, M., Ouwerkerk, P.B., Rueb, S., Delseny, M., and Guiderdoni, E. (2003). Highly efficient production and characterization of T-DNA plants for rice (Oryza sativa L.) functional genomics. Theor. Appl. Genet. **106:** 1396–1408.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, NY.

Thomas, C.M., and Jones, J.D. (2007). Molecular analysis of Agrobacterium T-DNA integration in tomato reveals a role for left border sequence homology in most integration events. Mol. Genet. Genomics **278:** 411–420.

Topping, J.F., Agyeman, F., Henricot, B., and Lindsey, K. (1994). Identification of molecular markers of embryogenesis in *Arabidopsis thaliana* by promoter trapping. Plant J. **5:** 895–903.

Topping, J.F., and Lindsey, K. (1997). Promoter trap markers differentiate structural and positional components of polar development in *Arabidopsis*. Plant Cell **9:** 1713–1725.

van der Graaff, E., Dulk-Ras, A.D., Hooykaas, P.J., and Keller, B. (2000). Activation tagging of the *LEAFY PETIOLE* gene affects leaf petiole development in *Arabidopsis thaliana*. Development **127:** 4971–4980.

van der Graaff, E., Hooykaas, P.J., and Keller, B. (2002). Activation tagging of the two closely linked genes LEP and VAS independently affects vascular cell number. Plant J. **32:** 819–830.

Wan, S., Wu, J., Zhang, Z., Sun, X., Lv, Y., Gao, C., Ning, Y., Ma, J., Guo, Y., Zhang, Q., Zheng, X., Zhang, C., Ma, Z., and Lu, T. (2009). Activation tagging, an efficient tool for functional analysis of the rice genome. Plant Mol. Biol. **69:** 69–80.

Wei, W., Twell, D., and Lindsey, K. (1997). A novel nucleic acid helicase gene identified by promoter trapping in *Arabidopsis*. Plant J. **11:** 1307–1314.

Wenck, A., Czako, M., Kanevski, I., and Marton, L. (1997). Frequent collinear long transfer of DNA inclusive of the whole binary vector during Agrobacterium-mediated transformation. Plant Mol. Biol. **34:** 913–922.

Wu, C., Li, X., Yuan, W., Chen, G., Kilian, A., Li, J., Xu, C., Li, X., Zhou, D.X., Wang, S., and Zhang, Q. (2003). Development of enhancer trap lines for functional analysis of the rice genome. Plant J. **35:** 418–427.

Wu, Y.R., Wang, Q.Y., Ma, Y.M., and Chu, C.C. (2004). Isolation and expression analysis of salt up-regulated ESTs in upland rice using PCR-based substractive suppression hybridization method. Plant Sci. **168:** 847–853

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y.,

Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., and Yang, H. (2002). A draft sequence of the rice genome *(Oryza sativa L. ssp. indica)*. Science **296:** 79–92.

Zhang, J., Guo, D., Chang, Y., You, C., Li, X., Dai, X., Weng, Q., Zhang, J., Chen, G., Li, X., Liu, H., Han, B., Zhang, Q., and Wu, C. (2007). Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by analyzing 13 804 T-DNA flanking sequences from an enhancer-trap mutant library. Plant J. **49:** 947–959.

Zhang, J., Li, C., Wu, C., Xiong, L., Chen, G., Zhang, Q., and Wang, S. (2006). RMD: a rice mutant database for functional analysis of the rice genome. Nucleic Acids Res. **34:** D745–D748.